



ISSN: 0975-766X
CODEN: IJPTFI
Research Article

Available Online through
www.ijptonline.com

EFFECTIVE PROCESSING OF QUESTION GENERATION FROM TEXT

Sunil Sagar, Prof. SenthilKumar N

School of Information Technology & Engineering, Vellore Institute of Technology, Vellore, Tamilnadu, India.

Email: senthilkumar.n@vit.ac.in

Received on 25-10-2016

Accepted on 02-11-2016

Abstract

Large number of text available on internet for educational purpose (e.g., news, Wikipedia, encyclopedia). However, to use this text in schools and colleges for educational purpose, there is many challenges exist, some of which is they don't have exercises or practice papers etc. Here, we try to implement a system for automating solves this problem by generating questions. We primly focus on generating WH type question using input sentence. Goal is to generate questions that takeinput sentences and generate output as questions using information available in the given input sentences by end user. First we introduce the problem, and then describe challenges available in question generation of linguistic and computational. After that we present a system for question generation that using the existing natural language tools and application for helps in question generation. System uses the manually written rules for transformation and processing some steps. Some experiments done on question generation and that result gives good question that are valuable. Natural language processing may help students and teachers to generate questions.

Keywords: Question Generation, Parser, Users, Automatic.

1. Introduction

What you say if someone ask, you have read or remember sentence or text? They may tell you to give explanation or describe. May tell you to describe its weaknesses and strengths. They might ask some questions to check your basic knowledge and you remember basic things. And they might ask challenging questions. If you are a good reader then you can give answer related to questions of article. However, lots of reader not good as much as others. So, generating manually. Questions are more time consuming and slow process. Wetry to automating this process of question generation. We try to generate question from sentences that is usable. The goal is to generate questions from given text, and created number of good questions. Text may be a web page or Wikipedia article or anything meaningful data. A user

then uses this question for practice or to know about article or use for quiz or test and gain knowledge about article topic.

We primarily focus on generating informational question and opinion questions. We have not considered basic use of unspecific area related knowledge of any topic, (e.g. information about history, geographical position). We try to generate fairly good question using lexical and syntactic process. The useful types are provided by much research paper. They described how questions can be arranged in many ways according to their characteristics, question purpose, the information question gives, question source of information, question answer length. What the purposes is of questions that system generating. There are many type of purposes: the knowledge based purpose (e.g., information related questions such as How many people are there?), common questions (e.g., a teacher ask student, what is the difference between boy and girl?). Generate question, what type of information they are containing. Question that contain very complex answers or simple questions. But it's requiring a well-trained system to generate complex good questions from input. This system generates many type of information related question such as, concept related question that contain particular information (e.g. who is the first prime minister of India?). Questions of yes and no type (e.g. was modi prime minister of India.). goal related question (e.g., why modi support other parties?). The source of information is important for question generated by system. Human generated question may base on common sense, world knowledge but system generated question only based upon the sentence or paragraph not on common sensor world knowledge. The answer phrase of question is short because it's not easy to work on long answer using parse tree so maximum answer length is one word. Question generation is helpful for teachers and students. It's generating deeper meaning question as well as simple question. Deeper question help to understand information batter. The important thing is that while teachers capable of generating any type of question then what is the need of the system for generating question for teacher. Well, this system is very useful somehow, my goal is to generate questions, and help teacher in generating questions in minimum time so it can save time and well as mental load.

2. Related Work

Past research has work on sentence level and generated information based question from sentence level data. Other techniques [2] of Question generation also generate question by sentence extraction and transformation into parse tree as I did in first step. And many question generation research uses English language for research as we did. Existing question generation system [3] is different in several ways. First of all they have used different rules for sentence extraction and

question generation, for example, they have used separate rules for who question that extract from main clauses, and for where question phrase also different rule to extract from main clauses. Then different rules for question generation from relative clauses. Also, existing work generate questions in single stem, they simply convert sentence into question using information extraction from clauses. Here, We generate question by applying many general rules[7]. The question generation process work in multi-step: first sentences extracted from the given input, and then working sentences on sentences generate question by using sequence of steps, like subject –auxiliary inversion and WH phrase generation. Also,the prior work only focus on generation question, but not focus on problems in generated question or generate right question. We have discussed challenges next chapters. Most of the prior work of question generation is focused on specific domain like a part of English linguistics [6] so that question can be generated using general rules. Here, we are working on complete English linguistics data and generate possible number of questions.

3. Definitions and Conventions

“Text enters by end user” refer to end user who enter document or number of lines text E.g. Delhi is capital of India. “Answer part” term is possible answer of generating question E.g. Delhi. “Question Part” term use for generated questions E.g. What in What is capital of India? Structure of given text is represent by phrase tree, that contains POS (Part of Speech). Sentence generate question using Tregex, a language that work on query and Tsurgeon for tree manipulation.

4. Materials and Methodology

Stage 1 of system, the given text or sentence subdivide into many sub sentences and then simplify given sentence and perform lexical, syntactic and semantics operation. Our proposed system resolve pronoun problem using ARK ref tool. Stage 2, the given sentence convert to number of question by executing number of operation on sentence (WH-word insertion, sub. – auxiliary inversion, etc.).

4.1 Stage 1: Transformations of Input Data

Transform given input sentence into the simpler form of sentences and the convert into the questions. Extract simple sentence from complex sentences by two ways: first we divide the given input sentence into number of sub sentence according to conjunction and full stops. Second step is to resolve pronoun to noun and then make sun sentences of input

sentence. Sentences give many information so in my implementation we take the complex sentence and the extract

simple sentences from the complex sentence and the generate questions according to sentence.

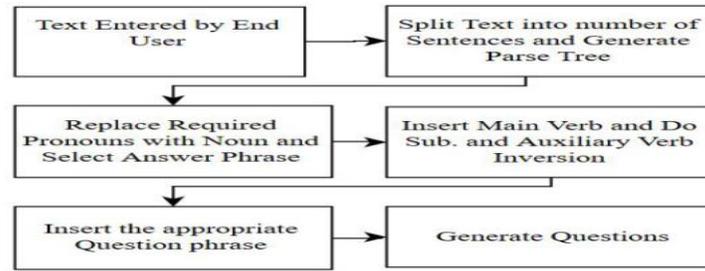


Figure 1: Process of Question Generation.

According to semantic entailment [4] if X entails Y means if X is true then Y is also true. This section show how we do simplification from complex to simple sentences by removing the connectives modifiers and verb phrases. This technique produce the meaningful sentences from complex sentence and from question can be generated. In order to overcome this pronoun problem, were place the pronoun with proper noun according to the given input sentence. Most probably the noun for pronoun is find easily by just knowing the previously following noun and replace that pronoun with followed noun but it's not give always correct pronoun replacement, in some case it gives wrong noun for pronoun also [1].

4.2 Stage 2: Question Generation

Question Generation produce number of question from given sentence. It identifies the answer phrase and then generate question phrase according to the selected answer phrase [8]. From the answer phrases system generate following type of question words for answer phrases: what, who, when, where, and whose. System also can generate how, why type questions also because the question phrase generation from answer. it can also generate these questions.

There are two cases in question generation. First are yes-no answer type questions so this type questions don't have any answer phrase to remove and then generate question. This type of question generates directly using subject auxiliary inversion. Second type of question is WH questions. That question contains WH word as the question and in this type of sentence I don't need to do subject auxiliary inversion [5]. We have to just decompose main verb and then select WH phrase according to the answer phrase and then generate question for that. If answer phrase is subject of the sentence then the subject of the sentence remove and then subject replace by the question phrase according to the subject. And main verb is selected according to the subject of the sentence like if first or second pronoun.

4.2.1 Selecting Unmovable Phrases

System marks the unmovable phrase using the Tregex expressions [3]. It does contain the format if the format is found in the sentence then that part of the sentence is marked as the unmovable part. If Tregex expression found any word or sentence unmovable then its mark that word or sentence as unmovable by add “unmovable” prefix in the word of parse tree. Then, while generating questions these words and sentences are eliminated if find prefix unmovable before words.

4.2.2 Question Phrases Generation

After find the unmovable phrase system find the answer phrase like subject of the sentence or objected. and then generate possible number of question according to that noun, location etc. In this process answer phrase removed from the sentence and generate many question according to answer phrase [6].But this step is not use in yes-no answer type question because they don't have answer phrase. Answer phrase can be any type like noun phrase that labeled by “NP”, “PP”, or “SBAR”. For noun phrases question can be generated of any type, basically it's depending upon the noun. And for subordinate clauses system generate direct question using what as question phrase.

System decomposes the main verb into the sentence if main verb not available in the sentence. For that we first check that main verb available in the sentence or not. If main verb available in sentence then we don't need to decompose the main verb into sentence otherwise decompose the main verb in the do from according to the sentence. If sentence in present tense the door does other if sentence in past tense then did so on. To convert form of the verb or in appropriate form of sentence to generate right question [2]. First we extract all needed verbs from the parse tree. Now, after extraction we find the base form of verb or required form of verb by using word net dictionary. It's give the required form of verb and if it's can't find the required verb form then it's use trimming method. In this method it's find verb base form by matching first one to one character because English verb form are generally made by adding in end extra postfix so it's easy to find base form. After answer phrase selection, system removes the answer phrase from parse tree and inserts the question phrases into the tree. So for one answer many type of question can be generated so it's made a separate copy for each question and save into separate parse tree [3].

For prepositional phrase answers, system generate the question depend upon the preposition like where or when questions. While post processing system removes the extra white spaces and adds the question mark at the end of question and removes the periods that come at the end of sentences for proper formatting of the question. System also

filter out the vague questions like questions that pronoun is not replaced by right noun or the question length more than the maximum limit of question generation length but such type of questions is very rare this occur few time in a while.

5. Results and Discussion

Sub sentence extraction from sentences generates more question and maximum sub sentences are right. But in some cases system generate wrong sub sentence because of wrong parse tree is generated. Label sentences with appropriate entity recognitions necessary to generate good question.

Named entity recognizer recognize the basic labels like person, location and origination but Supersence Tagger [1][7] is more usable because it's do all things of NER as well as find extra labels. Its use the part of speech tagging and near words feature for labeling and use hidden markov model. System generates many good questions as well as shows the answer of each question.

6. Conclusions and Future Work

We have worked only on one part of an automatic question generation there are many areas remaining to work on. Like selection of only hundred present right questions. Use other good approaches if available and generate more efficient and good questions. To generate question, we first convert input sentence into parse tree but there are may be other approaches are also usable. So we discuss here some approaches and the advantages and disadvantages of that approaches. There are still many challenges available in question generation and still need to solve all those challenges. Some of the challenges are related to computer linguistic problems and NLP problems. And challenges related to right question phrase generations also need research to improve good question phrase generation. Question generation system generates many questions but a higher rate of questions is vague. So to identify and resolve this problem that takes answers out of context and generating vague question we need more research on this. System also needs to generate deeper and more quality questions and for that I need more research on this area also. To generate deeper question system have to understand all data and work like common sense mean extract deep information.

References

1. Mitkov, R. and Ha, L. A. (2003). Computer-aided generation of multiple-choice tests. In Proc. Of the HLT-NAACL workshop on Building educational applications using natural language processing.

2. Kunichika, H., Katayama, T., Hirashima, T., and Takeuchi, A. (2004). Automated question generation methods for intelligent English learning systems and its evaluation. In Proc. Of ICCE.
3. Gates, D. M. (2008). Generating reading comprehension look-back strategy questions from expository texts. Master's thesis, Carnegie Mellon University.
4. Rus, V., Wyse, B., Piwek, P., Lintean, M., and Stoyanchev, S. (2010). Overview of the first question generation shared task evaluation challenge. In Proc. of the Third Workshop on Question Generation.
5. Mannem, P., Prasad, R., and Joshi, A. (2010). Question generation from paragraphs at UPenn: QG- STEC system description. In Proc. of the Third Workshop on Question Generation.
6. Yao, X. and Zhang, Y. (2010). Question generation with minimal recursion semantics. In Proc. of the Third Workshop on Question Generation.
7. Beigman Klebanov, B., Knight, K., and Marcu, D. (2004). Text simplification for information seeking applications. On the Move to Meaningful Internet Systems.
8. Carlson, A., Gaffney, S., and Vasile, F. (2009). Learning a named entity tagger from gazetteers with the partial perceptron. In Proc. of the AAAI Spring Symposium on Learning by Reading and Learning to Read.